

ЭМОЦИОНАЛЬНЫЕ ИНТЕЛЛЕКТУАЛЬНЫЕ АГЕНТЫ: СОХРАНЕНИЕ И ОБНОВЛЕНИЕ ГУМАНИЗМА

А. А. Карташева

Екатеринбург, Уральский федеральный университет

Аннотация. Способность к пониманию и управлению эмоциями означает, что человек может распознавать эмоции, контролировать их интенсивность и внешнее выражение. Концепт «эмоциональный искусственный интеллект» говорит о теоретической способности интеллектуальных агентов к распознаванию эмоций. Вопрос об эмоциональности интеллектуальных агентов может быть поставлен в двух плоскостях: с одной стороны, как вопрос об эмоциональной архитектуре самих агентов, а с другой — как возможность индуцирования эмоций в окружающих людях. В связи с этим неизбежно возникают изменения в оценке гуманистического потенциала интеллектуальных агентов.

Ключевые слова: эмоциональный интеллект, интеллектуальные агенты, гуманизм, распознавание эмоций

EMOTIONAL ARTIFICIAL INTELLIGENCE: PRESERVATION AND RENEWAL OF HUMANISM

Anna Kartasheva

Ekaterinburg, Ural Federal University

Abstract. The ability to understand and manage emotions means that the man can recognize emotions, control their intensity and external expression. The concept of “emotional artificial intelligence” refers to the theoretical ability of intelligent agents to recognize emotions. Question about emotional intelligent agents can be delivered in two planes: on the one hand, the question about the emotional architecture of agents themselves, and, on the other hand, the possibility of inducing emotions in the people around, in this context inevitably changes occur in the assessment of humanistic potential of intelligent agents.

Key words: emotional intelligence, intelligent agents, humanism, emotion recognition

Насколько люди эмпатичны? Правомерно ли говорить, что эмпатия свойственна всем людям без исключения? Вряд ли подобное утверждение является верным. Эмпатия каждого человека ограничена его личным опытом и психоэмоциональными реакциями, а также приобретаемыми стереотипами. Вполне вероятно, что наличие «эмпатичности» в архитектуре интеллектуальных агентов может быть сконструировано при помощи различных концепций, в различных степенях и с различными результатами.

Кроме того, вопрос об эмоциональности интеллектуальных агентов ставится в двух плоскостях. С одной стороны, интеллектуальные агенты способны быть эмоциональными, с другой — они могут вызывать эмоции в окружающих людях. В последнем случае особенно значителен гуманистический потенциал развития эмоциональной сферы применительно к интеллектуальным агентам.

Не претендуя на полный обзор концепций моделирования эмоций и на всеобъемлющую оценку их гуманистического потенциала, попробуем в данной статье кратко очертить круг проблем, которые возникают при размышлении об эмоциональности интеллектуальных агентов.

Начнем с того, что «интеллектуальный агент» — понятие из такой области академических исследований, как «искусственный интеллект». В. К. Финн пишет, что «под искусственным интеллектом (ИИ) следует понимать компьютерные реализации средств точной эпистемологии — языков представления знаний и логик рассуждений» [1, с. 220]. Формализация рассуждений и языки представления знаний являются предметом изучения ИИ, а продукт этого научного направления — интеллектуальные системы, человекомашинные системы или интеллектуальные агенты. Таким образом, одним из значений (но не единственным!) понятия «интеллектуальный агент» является фиксация искусственной субъектности — представление об интеллектуальном агенте как некой автономной сущности.

Стоит обратить внимание, что люди часто воспринимают эмоциональность интеллектуальных агентов как угрозу собственному гуманистическому мироощущению. Выражением этого, в частности, является эффект «Зловещей долины». Японский инженер М. Мори предложил гипотезу, согласно которой роботы, действующие

достаточно похоже на настоящих людей, вызывают неприязнь и отвращение у живых наблюдателей. Инженер опросил последних об их чувствах при знакомстве с линейкой искусственных объектов, расположенных по степени их похожести на человека. Согласно данным М. Мори [2, р. 98], в начале опроса роботы назывались тем более симпатичными, чем они были больше похожи на настоящих людей. Но это работало до определенного предела: в какой-то момент человекоподобные роботы стали вызывать чувство дискомфорта и страха. И на графике симпатий появлялось длинное плато — «Зловещая долина».

Кроме того, известна так называемая «проблема Элизы». В 1965 г. Д. Вейценбаум представил компьютерный симулятор человеческого общения, который назвал ELIZA. Концепция программы, по словам Вейценбаума [3], была реализована по модели беседы пациента с психотерапевтом: вопросы психотерапевта пациент истолковывает в зависимости от собственного душевного состояния, так как предполагает, что в вопросах психотерапевта изначально должен быть заложен смысл. В ходе тестирования программы Д. Вейценбаум наблюдал феномен, который позднее был назван «эффектом ELIZA», то есть склонностью человека проецировать свои чувства на машины. Д. Хофштадтер пишет о гораздо большей восприимчивости людей к содержанию сообщений от компьютеров, чем это необходимо и оправданно [4]. Желание людей полагать, что программа устроена сложнее, чем это есть в реальности, вызывало тревогу Д. Вейценбаума относительно дальнейшего существования социальности и человечности [5]. Напротив, М. Минский утверждал [6], что вполне достижимо (и даже необходимо) воплощение нейробиологической структуры человеческого мышления в вычислительной архитектуре компьютеров.

Давайте предположим, что развитие эмоциональности интеллектуальных агентов не несет угрозы гуманизму, а, наоборот, служит его сохранению и обновлению. Гипотеза о том, что эмоциональность интеллектуальных агентов не будет негативно влиять на гуманистические ценности, помогает подойти к преодолению эффекта «Зловещей долины» с точки зрения рационального осмысления.

Как это возможно? Во-первых, моделирование эмоциональности интеллектуальных агентов не заменяет эмоциональности человеческой: речь о человекомашинном взаимодействии. Во-вторых, вспомним про две плоскости рассмотрения эмоциональности у интеллектуальных агентов. Возможность интеллектуальных систем индуцировать эмоции в окружающих людях безусловно несет в себе гуманистический потенциал, так как вызываемые эмоции чаще всего являются положительными.

Программы исследований в области человекомашинного взаимодействия (HCI) обязательно включают в себя эмоциональное взаимодействие (*affective interaction*). Так, Р. Пикар говорит об эмоциях, создаваемых в коммуникации — как между людьми, так между людьми и машинами [7]. Исследования эмоционального взаимодействия опираются на феноменологию как философское направление и на работы в области HCI. Если согласиться с Р. Пикар, что эмоции создаются во взаимодействии, то их моделирование служит сохранению гуманизма, а также его обновлению при условии соблюдения высоких моральных нравственных стандартов при моделировании эмоций у интеллектуальных агентов. Для этого необходим общественный контроль и, возможно, открытость процессов моделирования вычислительных архитектур.

Исследования в области человекомашинного взаимодействия актуальны и для сервисных роботов, которые могут заменить человека при выполнении отдельных работ и которые отличаются от промышленных степенью близости к конечному пользователю. О. И. Муштак пишет, что «у социальных роботов (которых можно условно отнести к подвиду сервисных роботов, хотя исследователи все еще не имеют точных оценок касательно соотношения этих терминов) степень этой близости максимальная, пользователь не исключается из процесса работы робота, не выступает в роли „наблюдателя“ над машиной, а взаимодействует с ней привычными человеку способами» [8, с. 72]. Необходимость коммуникации людей с роботами открывает перспективы для изучения способов, методов и нюансов взаимодействия.

В рамках данной статьи не поднимается актуальный вопрос об эпистемических ограничениях в рамках анализа взаимодействий

людей и интеллектуальных агентов, а также не представлен спектр трактовки понятий «взаимодействие» и «агентность». Заинтересованного читателя я хотела бы переадресовать к статье Е. А. Гаврилиной [9], где подробно рассматриваются теоретические ограничения, которые могут искажать эмпирические результаты HCI-исследований.

Будет ли взаимодействие носить двусторонний характер или представлять собой однонаправленное индуцирование эмоций в окружающих людях, важно теоретически обосновать и оценивать не один, а сразу несколько параметров. Необходимость синхронизации внешности и поведения антропоморфных роботов подчеркивается в исследовании Е. В. Середкиной, И. И. Безукладникова и М. С. Долгих. Исследователи отмечают, что «избыточность антропоморфной формы при низком поведенческом функционале может стать барьером для эффективного взаимодействия между людьми и роботами» [10, с. 226]. Человек ожидает общения, а робот, внешне похожий на человека, не может дать паттернов поведения, типичных для человека.

Итак, развитие интеллектуальных агентов может восприниматься людьми в качестве угрозы гуманизму (в частности, значимы эффект «Зловещей долины» и «эффект ELIZA»). Если посмотреть на эту проблему с точки зрения гуманистического мироощущения, то есть несколько путей дальнейших исследований для сохранения и обновления гуманизма. Так, изучение эмоциональности интеллектуальных агентов затрагивает эмоциональное взаимодействие в рамках человекомашинного взаимодействия. Или же в будущем могут исследоваться возможности интеллектуальных агентов в одностороннем порядке индуцировать эмоции в окружающих людях.

ЛИТЕРАТУРА

1. Финн В. К., Михеенкова М. А. Точная эпистемология, искусственный интеллект и интеллектуальный анализ данных в науках о жизни и социальном поведении // Проектирование будущего. Проблемы цифровой реальности : тр. 3-й Международ. конф. (Москва, 6–7 февраля 2020 г.). М. : ИПМ им. М. В. Келдыша, 2020. С. 218–226. URL: <https://doi.org/10.20948/future-2020-19> (дата обращения: 10.10.2020).

2. *Mori M.* The Uncanny Valley // IEEE Robotics & Automation Magazine. 2012. Vol. 19, № 2. P. 98–100. DOI:10.1109/MRA.2012.2192811 (date of access: 10.10.2020).
3. Документальный фильм WEIZENBAUM Rebel at Work. URL: <https://www.youtube.com/watch?v=hAfpnkaZprE> (дата обращения: 10.10.2020).
4. *Hofstadter D.* Fluid Concepts and Creative Analogies: Computer Models of the Fundamental Mechanisms of Thought. N. Y. : Basic Books, 1995.
5. *Weizenbaum J.* Computer Power And Human Reason. From Judgment to Calculation // The Massachusetts Institute of Technology. N. Y. : W. H. Freeman and Company, 1976.
6. *Minsky M.* The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind. N. Y. : Simon & Schuster, 2006.
7. *Picard R.* Affective Computing. Cambridge, Massachusetts : MIT Press, 1997.
8. *Муштак О. И.* Гуманистический искусственный интеллект: эмоциональные аспекты человекомашинного взаимодействия // Гуманизация современной философии и гуманитаризация общественных наук : материалы науч. конф. (Екатеринбург, 17–18 мая 2018 г.). Екатеринбург : Деловая книга, 2018. С. 71–74. URL: <http://hdl.handle.net/10995/62918> (дата обращения: 10.10.2020).
9. *Гаврилина Е. А.* Как люди взаимодействуют с роботами: между теоретическими поисками и эмпирическими исследованиями // Технологос. 2020. № 2. С. 5–14. DOI: 10.15593/perm.kipf/2020.2.01 (дата обращения: 10.10.2020).
10. *Середкина Е. В., Безукладников И. И., Долгих М. С.* Прикладной антропоморфизм в сервисной робототехнике // Интеллектуальные системы в науке и технике. Искусственный интеллект в решении актуальных социальных и экономических проблем XXI века : сб. ст. по материалам Международ. конф. «Интеллектуальные системы в науке и технике» и Шестой Всерос. науч.-практ. конф. «Искусственный интеллект в решении актуальных социальных и экономических проблем XXI века» (Пермь, 12–18 октября 2020 г.) / под ред. Л. Н. Ясницкого ; Перм. гос. нац. исслед. ун-т. Пермь, 2020. URL: <http://www.psu.ru/files/docs/science/books/sborniki/iskusstv-intellekt-v-reshenii-akt-problem-xxi-veka.pdf>. С. 223–227 (дата обращения: 10.10.2020).